How to use Variant Effects Report

A. Introduction to Ensembl Variant Effect Predictor
B. Using RefSeq_v1
C. Using TGACv1

A. Introduction

The Ensembl Variant Effect Predictor is a toolset for the analysis, annotation, and prioritization of genomic variants in coding and non-coding regions. There are more than 1 million single nucleotide variants in wheat. There may be only 10 thousand that change the amino acid coding and a smaller subset of these that truncate or produce a loss of function.

The Ensembl Variant Effect Predictor (VEP) can be accessed by web, Perl script, web based API. The report page shows calculations for markers in T3 and provides links to calculation provided by Ensembl Plant

The inputs used for the VEP are

a. Ensembl or VCF format SNP locations
b. High confidence (HC) gene predictions
c. FASTA file of the assembly

The outputs of the VEP are

a. Feature – transcript (can also include motif and regulatory elements)
b. Consequence
c. Impact

A detailed description of the Consequence and Impact values can be found here:
http://ensembl.org/info/genome/variation/predicted_data.html

The Ensembl VEP also incorporates SIFT and PolyPhen-2 but these are not used on the T3 website.

Sorting Intolerant From Tolerant (SIFT) predicts whether an amino acid substitution is likely to affect protein function based on sequence homology and the physico-chemical similarity between the alternate amino acids.

SIFT value	Qualitative prediction	Website display example	
Less than 0.05	"Deleterious"	0.01	0.01
Greater than or equal to 0.05	"Tolerated"	0.8	0.8

PolyPhen-2 predicts the effect of an amino acid substitution on the structure and function of a protein using sequence homology, Pfam annotations, 3D structures from PDB where available, and a number of other databases and tools (including DSSP, ncoils etc.)

Consequence values for VEP

3_prime_UTR_variant
5_prime_UTR_variant
coding_sequence_variant
coding_sequence_variant,3_prime_UTR_variant
coding_sequence_variant,5_prime_UTR_variant
downstream_gene_variant
frameshift_variant
frameshift_variant,splice_region_variant
frameshift_variant,splice_region_variant,intron_variant
frameshift_variant,start_lost
frameshift_variant,start_lost,splice_region_variant
frameshift_variant,start_lost,start_retained_variant
frameshift_variant,stop_lost
frameshift_variant,stop_lost,splice_region_variant
frameshift_variant,stop_retained_variant
inframe_deletion
inframe_deletion,splice_region_variant
inframe_insertion
inframe_insertion,splice_region_variant
inframe_insertion,stop_retained_variant
intergenic_variant
intron_variant
protein_altering_variant
protein_altering_variant,splice_region_variant
splice_acceptor_variant
splice_acceptor_variant,5_prime_UTR_variant
splice_acceptor_variant,coding_sequence_variant
splice_acceptor_variant,coding_sequence_variant,intron_variant
splice_acceptor_variant,frameshift_variant
splice_acceptor_variant,inframe_insertion
splice_acceptor_variant,intron_variant
splice_donor_variant
splice_donor_variant,coding_sequence_variant
splice_donor_variant,coding_sequence_variant,5_prime_UTR_variant
splice_donor_variant,coding_sequence_variant,intron_variant
splice_donor_variant,frameshift_variant
splice_donor_variant,inframe_insertion
splice_donor_variant,intron_variant
splice_region_variant,3_prime_UTR_variant
splice_region_variant,5_prime_UTR_variant
splice_region_variant,intron_variant
start_lost
start_lost,5_prime_UTR_variant
start_lost,inframe_deletion
start_retained_variant
start_retained_variant,5_prime_UTR_variant
stop_gained
stop_gained,frameshift_variant
stop_gained,frameshift_variant,splice_region_variant
stop_gained,inframe_deletion
stop_gained,inframe_insertion
stop_gained,inframe_insertion,splice_region_variant
stop_gained,splice_region_variant
stop_lost,3_prime_UTR_variant
stop_lost,inframe_deletion
stop_retained_variant,3_prime_UTR_variant
upstream_gene_variant

Impact values for VEP

HIGH - disruptive impact in the protein, protein truncation or loss of function
LOW – harmless, unlikely to change protein behaviour
MODERATE - non-disruptive variant that might change protein effectiveness
MODIFIER – usually non-coding variants

General instructions for to use the T3 VEP report

First select a list of markers (limit the selection to under 1000). It will accept a mix of markers from different genotype experiments or a single genotype experiment. The positions of the markers on the genome assembly have been identified either by BLAST or from the coordinates provided when the genotype results were loaded into the database. If the marker position cannot be identified then it will be listed at the bottom of the page as not found.

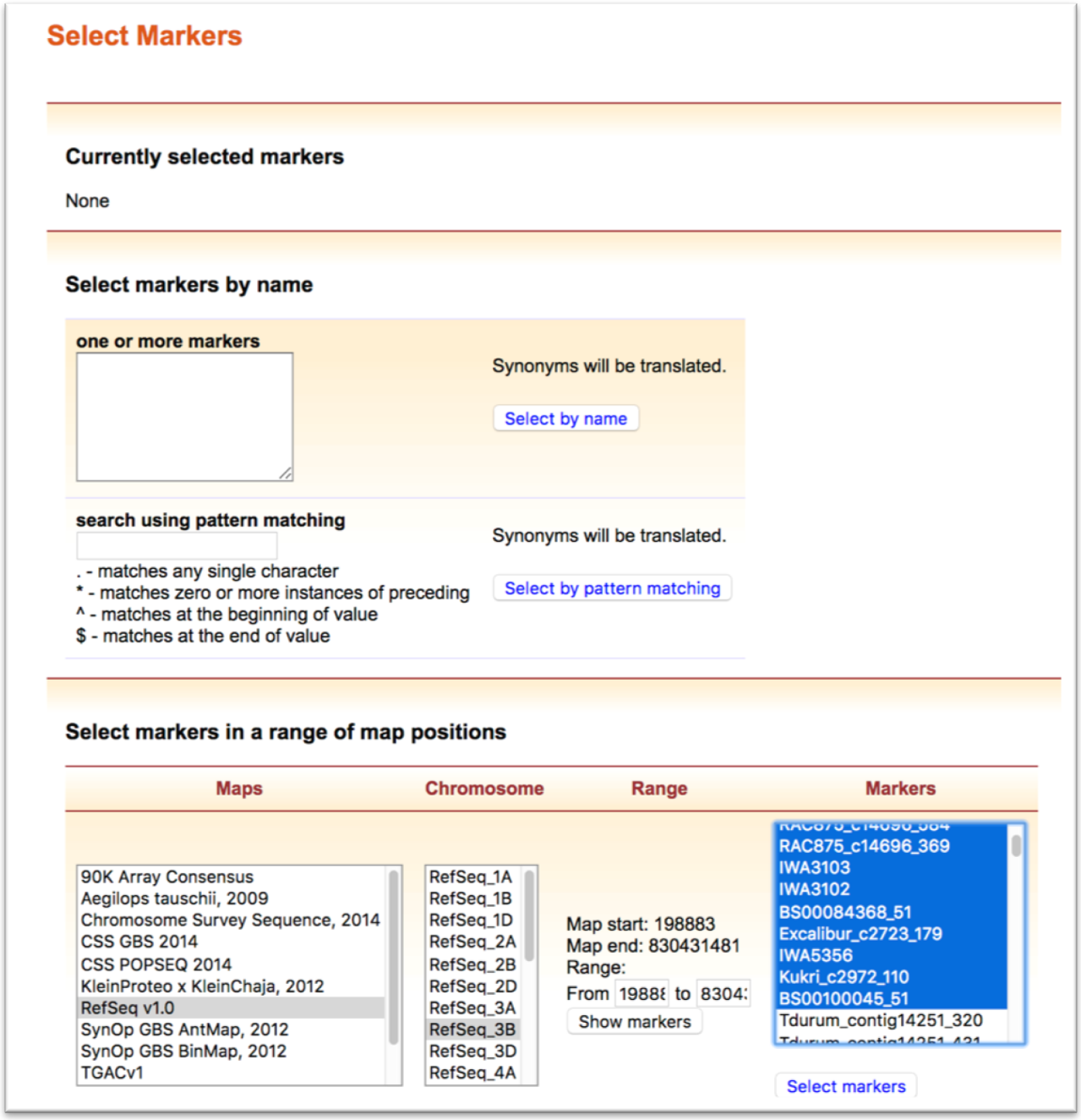
For markers not found on the map

You can run BLAST against RefSeq and format the output in either Ensembl or VCF format then mail the file to me using the feedback link on the T3 website. Then I can run the Ensembl VEP program on our machine and email you the results.

B. Using RefSeq_v1 assembly – the only markers that have been mapped to RefSeq_v1 are in the RefSeq v1.0 Physical Map and the 2017_WheatCAP genotype experiment.

1. Visit https://triticeaetoolbox.org/wheat/

2. Select markers of interest. Go to Select => Markers



3. View the Variant Effects: Go to Reports => Variant Effects

Variant Effects

This page provides links to Sorting Intolerant From Tolerant (SIFT) and Variant Effect Predictor (VEP) to predict whether an amino acid substitution affects protein function. SIFT missense predictions for genomes: Nature Protocols 2016; 11:1-9. The Ensembl Variant Effect Predictor: Genome Biology Jun 6;17(1):122. (2016) doi:10.1186/s13059-016-0974-4.

Genome Assembly RefSeq_v1

The links in the region column show known variations in a genome browser and their effects. The region is 1000 bases to either side of marker.
The links in the gene column show a table with known variations, consequence type, and SIFT score.

marker	region	gene	description	feature	consequence	impact
RAC875_c11409_550	chr3B:2958657	TraesCS3B01G005000	receptor kinase 1	-	intergenic_variant	MODIFIER
gbsCNLmaster28410	chr3B:2959834	TraesCS3B01G005000	receptor kinase 1	-	intergenic_variant	MODIFIER
gbsHWWAMP34169	chr3B:2959834	TraesCS3B01G005000	receptor kinase 1	-	intergenic_variant	MODIFIER
Excalibur_c3031_59	chr3B:3036848	TraesCS3B01G005100	receptor kinase 1	TraesCS3B01G005100.1	downstream_gene_variant	MODIFIER
gbsHWWAMP55562	chr3B:3042892	TraesCS3B01G005100	receptor kinase 1	TraesCS3B01G005100.1	upstream_gene_variant	MODIFIER
IACX4260	chr3B:3228129	TraesCS3B01G005800	Paired amphipathic helix SIN3-like protein	TraesCS3B01G005800.1	frameshift_variant	HIGH
gbsHWWAMP49959	chr3B:3228452	TraesCS3B01G005800	Paired amphipathic	TraesCS3B01G005800.1	frameshift_variant	HIGH

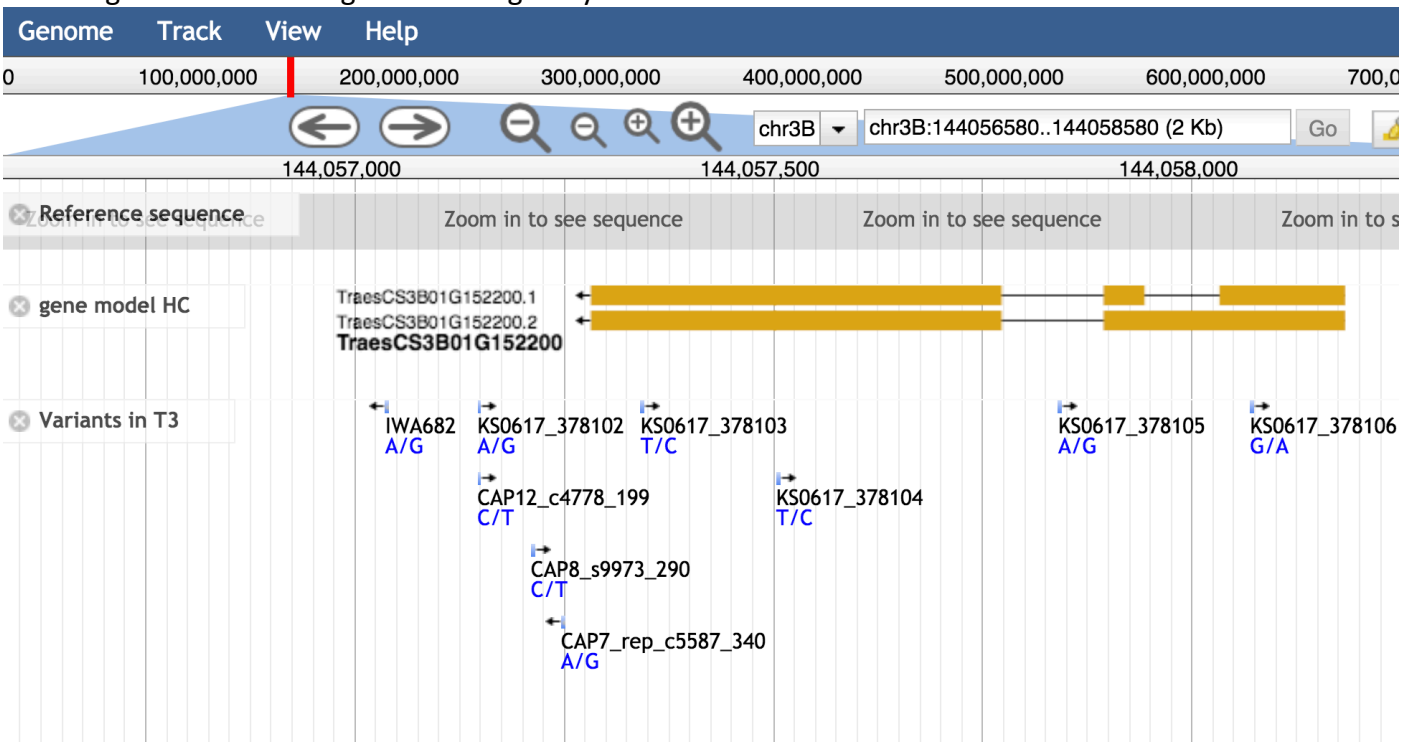
Selecting the link in the Gene column gives you a report of all markers for that gene.

Variant Effects Gene TraesCS3B01G005800

ENSEMBL VARIANT EFFECT PREDICTOR v90.4
Citation: McLaren et. al. 2016 (doi:10.1186/s13059-016-0974-4)

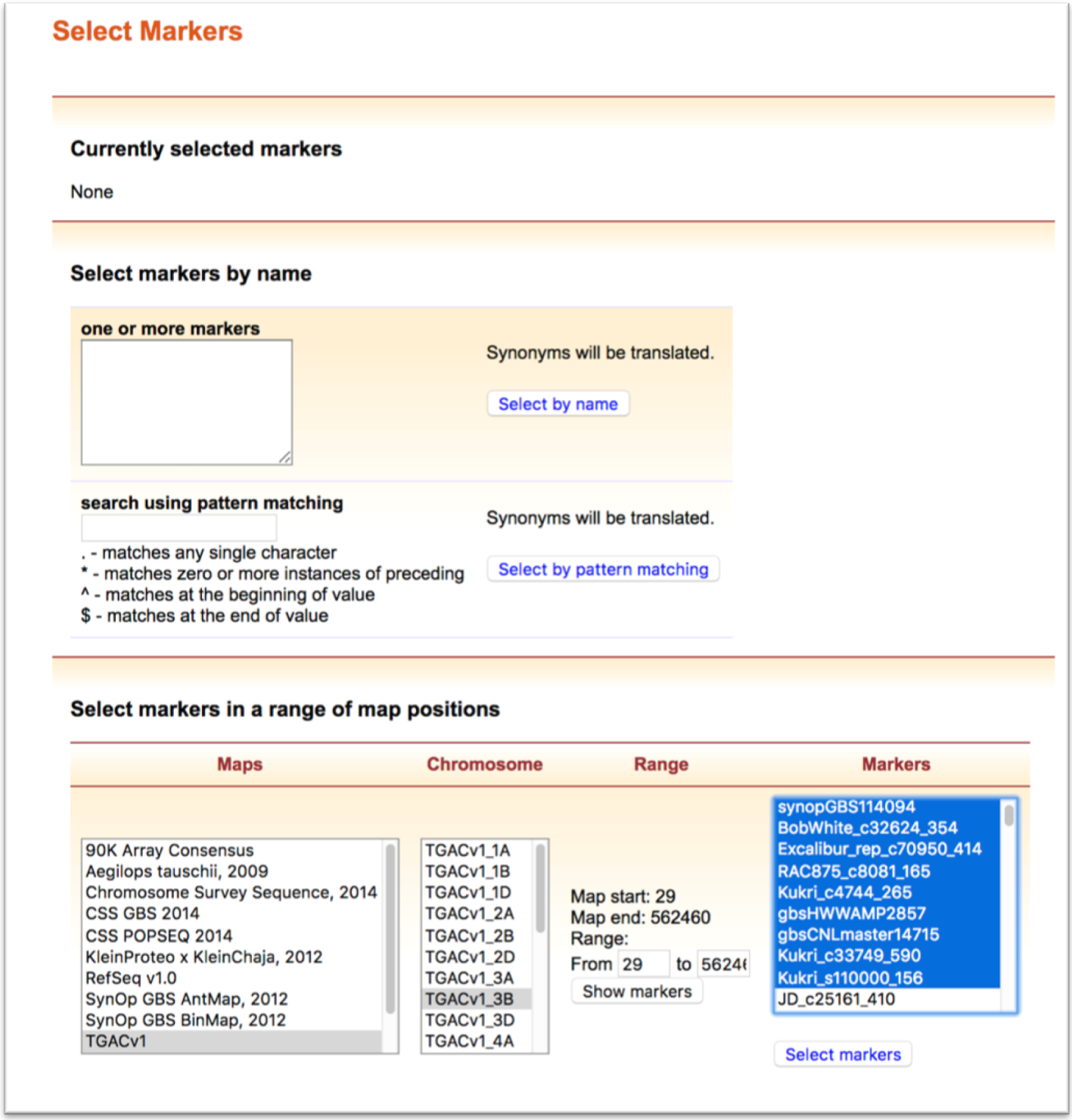
marker name	location	feature	consequence	impact
gbsHWWAMP49959	chr3B:3228452	TraesCS3B01G005800.1	frameshift_variant	HIGH
IACX4260	chr3B:3228129	TraesCS3B01G005800.1	frameshift_variant	HIGH
KS0617_361198	chr3B:3225869	TraesCS3B01G005800.1	5_prime_UTR_variant	MODIFIER
KS0617_361199	chr3B:3225892	TraesCS3B01G005800.1	splice_region_variant,5_prime_UTR_variant	LOW
KS0617_361200	chr3B:3225927	TraesCS3B01G005800.1	intron_variant	MODIFIER
KS0617_361201	chr3B:3225929	TraesCS3B01G005800.1	intron_variant	MODIFIER
KS0617_361202	chr3B:3225933	TraesCS3B01G005800.1	intron_variant	MODIFIER

Selecting the link in the region column gives you a JBrowse view for that marker

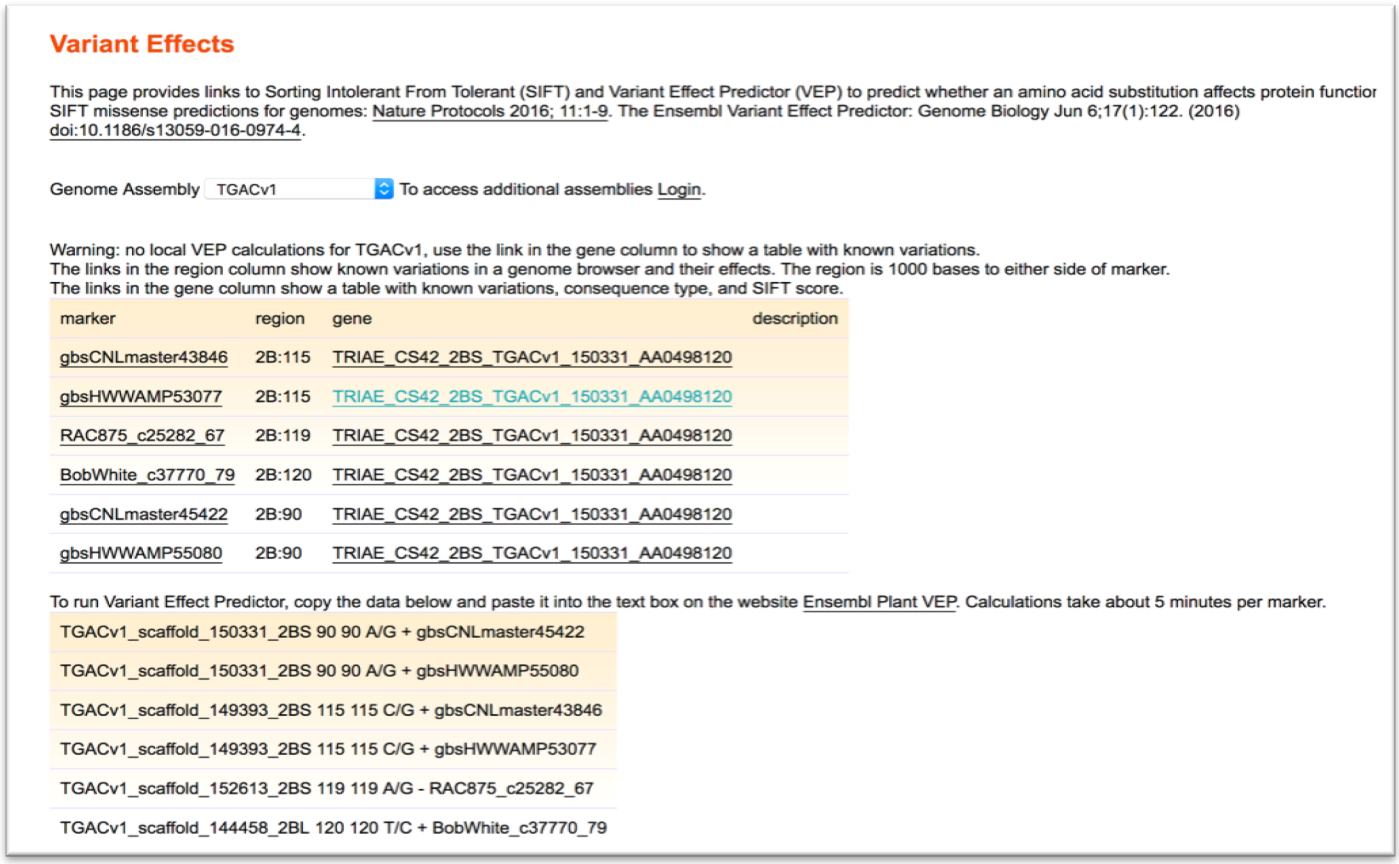


C. Using TGACv1 assembly

1. Select markers of interest. Go to Select => Markers



2. View the Variant Effects: Go to Reports => Variant Effects
3. Select TGACv1 for the Genome Assembly
4. To view the Variant Effect, you can either
    a. Click on the link in the Gene column. This will take you to a table on Ensembl Plant website.
    b. Scroll down the page until you see the second table. Copy these entries and past them in the data field of the Variant Effect Predictor on the Ensembl Plant website.



Variant Effects

This page provides links to Sorting Intolerant From Tolerant (SIFT) and Variant Effect Predictor (VEP) to predict whether an amino acid substitution affects protein functior
SIFT missense predictions for genomes: Nature Protocols 2016; 11:1-9. The Ensembl Variant Effect Predictor: Genome Biology Jun 6;17(1):122. (2016)
doi:10.1186/s13059-016-0974-4.

Genome Assembly TGACv1 To access additional assemblies Login.

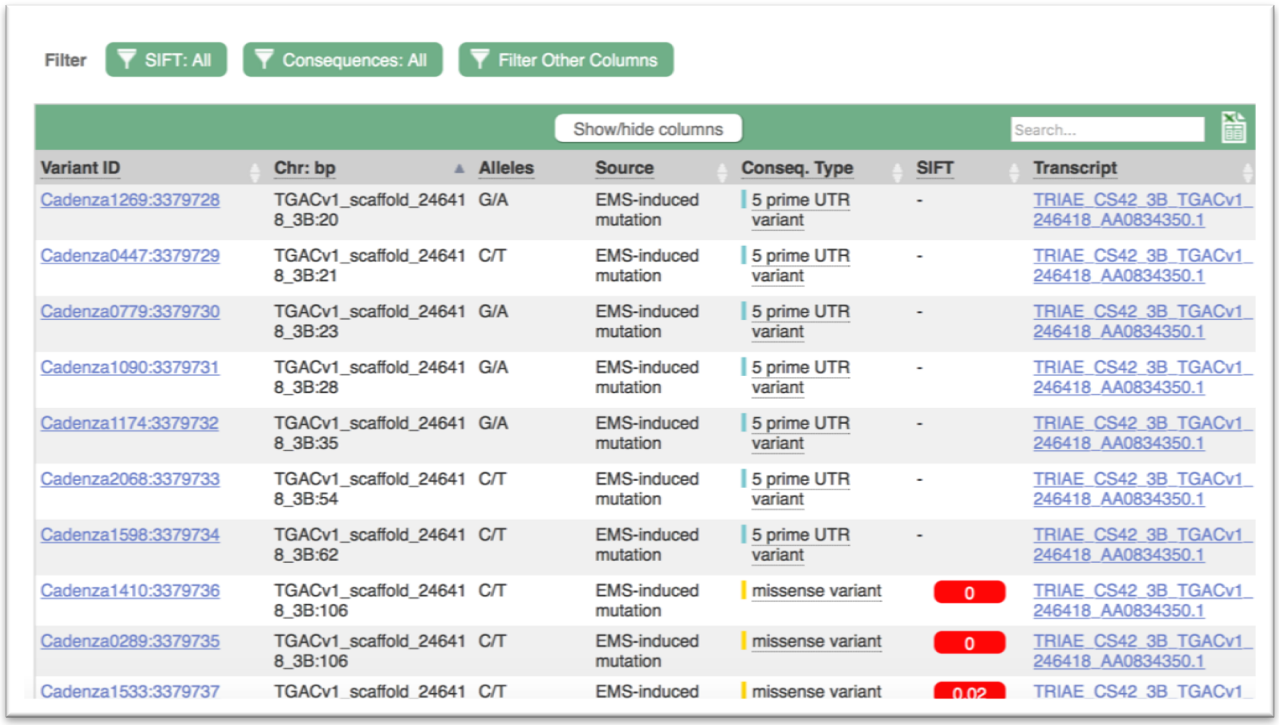
Warning: no local VEP calculations for TGACv1, use the link in the gene column to show a table with known variations.
The links in the region column show known variations in a genome browser and their effects. The region is 1000 bases to either side of marker.
The links in the gene column show a table with known variations, consequence type, and SIFT score.

marker	region	gene	description
gbsCNLmaster43846	2B:115	TRIAE_CS42_2BS_TGACv1_150331_AA0498120	
gbsHWWAMP53077	2B:115	TRIAE_CS42_2BS_TGACv1_150331_AA0498120	
RAC875_c25282_67	2B:119	TRIAE_CS42_2BS_TGACv1_150331_AA0498120	
BobWhite_c37770_79	2B:120	TRIAE_CS42_2BS_TGACv1_150331_AA0498120	
gbsCNLmaster45422	2B:90	TRIAE_CS42_2BS_TGACv1_150331_AA0498120	
gbsHWWAMP55080	2B:90	TRIAE_CS42_2BS_TGACv1_150331_AA0498120	

To run Variant Effect Predictor, copy the data below and paste it into the text box on the website Ensembl Plant VEP. Calculations take about 5 minutes per marker.

```
TGACv1_scaffold_150331_2BS 90 90 A/G + gbsCNLmaster45422
TGACv1_scaffold_150331_2BS 90 90 A/G + gbsHWWAMP55080
TGACv1_scaffold_149393_2BS 115 115 C/G + gbsCNLmaster43846
TGACv1_scaffold_149393_2BS 115 115 C/G + gbsHWWAMP53077
TGACv1_scaffold_152613_2BS 119 119 A/G - RAC875_c25282_67
TGACv1_scaffold_144458_2BL 120 120 T/C + BobWhite_c37770_79
```

Selecting the link in genes column directs you to Ensembl Plant to show you a table of variants for that gene.



Filter | SIFT: All | Consequences: All | Filter Other Columns

Show/hide columns | Search...

Variant ID	Chr: bp	Alleles	Source	Conseq. Type	SIFT	Transcript
Cadenza1269:3379728	TGACv1_scaffold_246418_3B:20	G/A	EMS-induced mutation	5 prime UTR variant	-	TRIAE_CS42_3B_TGACv1_246418_AA0834350.1
Cadenza0447:3379729	TGACv1_scaffold_246418_3B:21	C/T	EMS-induced mutation	5 prime UTR variant	-	TRIAE_CS42_3B_TGACv1_246418_AA0834350.1
Cadenza0779:3379730	TGACv1_scaffold_246418_3B:23	G/A	EMS-induced mutation	5 prime UTR variant	-	TRIAE_CS42_3B_TGACv1_246418_AA0834350.1
Cadenza1090:3379731	TGACv1_scaffold_246418_3B:28	G/A	EMS-induced mutation	5 prime UTR variant	-	TRIAE_CS42_3B_TGACv1_246418_AA0834350.1
Cadenza1174:3379732	TGACv1_scaffold_246418_3B:35	G/A	EMS-induced mutation	5 prime UTR variant	-	TRIAE_CS42_3B_TGACv1_246418_AA0834350.1
Cadenza2068:3379733	TGACv1_scaffold_246418_3B:54	C/T	EMS-induced mutation	5 prime UTR variant	-	TRIAE_CS42_3B_TGACv1_246418_AA0834350.1
Cadenza1598:3379734	TGACv1_scaffold_246418_3B:62	C/T	EMS-induced mutation	5 prime UTR variant	-	TRIAE_CS42_3B_TGACv1_246418_AA0834350.1
Cadenza1410:3379736	TGACv1_scaffold_246418_3B:106	C/T	EMS-induced mutation	missense variant	0	TRIAE_CS42_3B_TGACv1_246418_AA0834350.1
Cadenza0289:3379735	TGACv1_scaffold_246418_3B:106	C/T	EMS-induced mutation	missense variant	0	TRIAE_CS42_3B_TGACv1_246418_AA0834350.1
Cadenza1533:3379737	TGACv1_scaffold_246418_3B	C/T	EMS-induced	missense variant	0.02	TRIAE_CS42_3B_TGACv1

On the Ensembl page there is a link to view the location in the Ensembl Browser, which will show you the position and variant types for that gene

You can run the Ensembl VEP program for your own markers by copying the results from the bottom table of the T3 Variant Effects page and pasting it into the tool on the Ensembl Plant website. http://plants.ensembl.org/Triticum_aestivum/Tools/VEP?db=core